Learning Basic Statistics

A Handbook

for

Medical students

Volume I

Prepared for the Community Stream
Faculty of Medicine
Colombo

By

Ms. Kantha N. Lankatilake
Dr. Rohini S. de A. Seneviratne
Dr. Lalani Rajapakse
Prof. Dulitha N. Fernando

LEARNING STATISTICS

A Handbook
For
Medical Students

Volume - 1

CONTENTS

page number

BASIC STATISTICS

The self learning modules on basic statistics are presented as two volumes. This book is the first volume and consists of two lessons. Each lesson contains the learning objectives, subject material, worked examples, assignments; and is designed for self learning. In addition, learning is to be supplemented with fixed learning modules which will be referred at various points in the relevant sections.

Note that the learning material given is only a guideline and is not complete by itself. It is expected that students will undertake further reading for better understanding of the subject. Recommended reading is also given in this book. One should also attempt to answer the questions given. The objectives of the module is to enable a student to develop a competency in basic statistical methods in order that these can be applied to data and also understand data presented in research papers.

page number

INTRODUCTION

Statistics

There are two different meanings to the term "statistics".

i) It is used to indicate facts and figures of any kind: health statistics, business statistics, vital statistics etc. eg - number of live births per 1000

ii) It is also used to refer to a body of knowledge known as statistical methods developed for handling numerical data in general i.e. collecting, organizing, summarising and interpreting data.

Biostatistics

Statistics arising out of biological sciences, particularly from the field of medicine. Statistical methods when applied to biological sciences and medicine is referred to as Biostatistics.

Statistics in general consists of two subdivisions: **Descriptive statistics** and predictive or **inferential statistics**.

Descriptive statistics includes techniques of describing data in abbreviated, symbolic fashion. This could be done by: presenting the data as a diagram or table, or by using a number to represent the data such as an average or mean, standard deviation, etc.

Inferential statistics, makes use of a limited number of observations from a valid sample to predict or infer characteristics of an entire group or population.

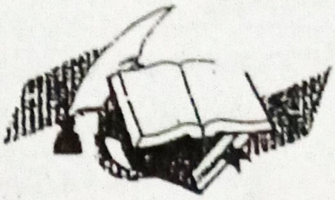
Measurement

A measurement is essentially the assigning of numbers to observations according to certain rules. All measurements are made according to one of four principal scales.

Scales of measurement

The four principal scales of measurement are:

1. Nominal scale
2. Ordinal scale
3. Interval scale
4. Ratio scale



categorical data

FURTHER READING

Swinscow, T.D.V. 1983. Statistics in Square One, London.

Sprinthall, R.C. Basic Statistical Analysis. (4th ed.) Allyn and Bacon.

Castle, W.M. 1977. Statistics in Small Doses. (2nd ed.). Churchill Livingstone, Edinburgh

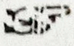
Petrie, A. 1978. Lecture Notes on Medical Statistics. Blackwell, Oxford.

Pipkin, B.F. 1984. Medical Statistics Made Easy. Churchill Livingstone, Edinburgh.

Teaching Health Statistics. 1986. Twenty Lessons and Seminar Outlines. Edited by S.K. Lwanga and Cho-Yook Tye. WHO, Geneva.

numerical data

SYMBOLS USED IN THE BOOKLET

 take note of

 exercises

 refer to Fixed Learning Module (FiLM)

 draw

Nominal scale (classificatory scale) - Observations are sorted or grouped categories. The number in each category is then counted e.g. students in your batch be put into four categories according to their blood groupings: A, B, AB, & number of students in each category could then be counted.

Ordinal scale (ranking scale) - Arranges observations into groups which form ordered series. Information regarding greater than or less than status is containe ordinal data, but information as to how much greater than or less than is not indic e.g. socio-economic status : lower class, middle class and upper class. Households community could be ranked according to their socio-economic status as low inc middle income and high income categories.

Interval scale - This is a scale in which the intervals between successive points equal. Measurements made using this scale contains information with regard to much greater than or how much less than a measurement is in comparison with ano e.g. measurements made in Fahrenheit (or Celsius) temperature scale.

Ratio scale - In this scale too, the intervals between successive points are equal. scale has an added advantage of having an absolute zero (as opposed to an arbitrary such as 0^0 Fahrenheit or Celsius temperature), which enables measurements to compared. e.g. a person 6 feet tall could be stated as being twice as tall as some who is 3 feet tall.

Variable - A variable is anything that can be observed to vary.

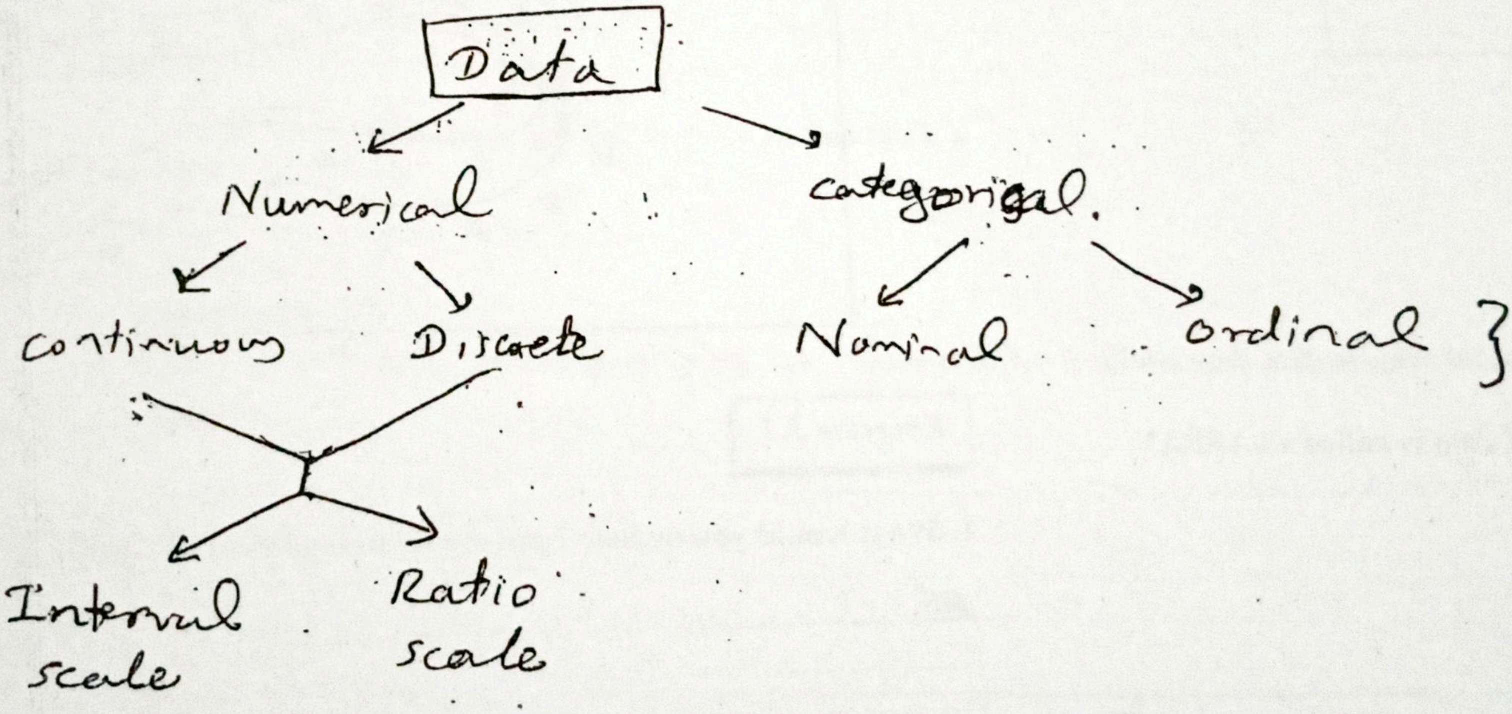
Variables could be categorised as continuous and discrete variables.

A **continuous variable** can take any value in a range of values. Heights in a sample medical students if found to range from 144.8cm to 172.7cm, means that a particu student's height can take any value within this range. It can either take a fractional va (172.724cm, 172.72cm, 172.7cm), or an integer or whole number (173cm) depending the degree of accuracy to which the measurements has to be made.

A discrete variable can only take a definite value in any given interval. Here the value of the variable is an integer e.g. parity : 1, 2, 3. Number of households in a community : 500, 525, 558 etc. This will not take a value such as 558.5

Variables could also be classified according to the scale of measurement as quantitative and qualitative.

Quantitative variables are variables measured using either the interval or ratio scale. e.g. height and weight. **Qualitative variables** are variables classified on the basis of some quality such as, sex, religion, ethnicity or nationality. This type of variable can be measured using only the nominal and ordinal scales. The measurement made refers to the number (frequency or count) in each category.



I - DESCRIPTIVE STATISTICS

LESSON 1

Non numerical methods of describing data

LEARNING OBJECTIVES

At the end of the lesson the student should be able to:

1. *Organize* the data set in an array and describe it using the minimum value, maximum value and range
2. Comment on the uses and limitations of such re-arranged data
3. *Summarize* the data into a frequency distribution, both ungrouped and grouped
4. Comment on the uses and limitations of summarized data
5. *Present* data by way of tabulations and diagrams
6. *Describe* data presented in a tabular or diagrammatic form
7. Comment on the advantages and disadvantages of such presentations

1. Here is a small data set

Figure 1.1

Number of deaths which occurred per day over a period of one month in hospital A.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Week 1		1	4	0	1	2	1
Week 2	3	3	0	2	0	1	3
Week 3	3	2	1	0	3	5	0
Week 4	2	4	2	1	3	2	1
Week 5	2	2	3				

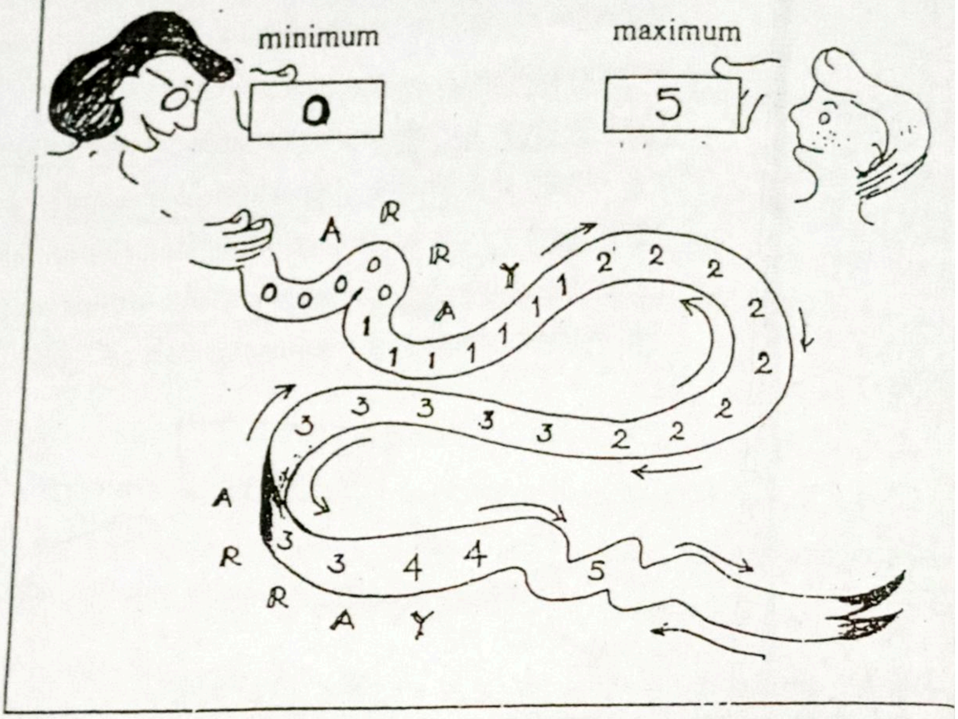
By examining the raw data, what comments can you make ?

☞ The raw data needs to be first organized in a meaningful way so that they could be easily understood. The very basic re-arrangement of data is called an ARRAY.

1.1 Steps in making an array

1. Find minimum and maximum values
2. Arrange the values either in an increasing or decreasing order of magnitude

Figure 1.2



Exercise 1.1

1. What would you deduce from the re-arrangement of data?

....
....
....
....

Well !! you could say that ....

- minimum number of deaths per day is 0.
- maximum number of deaths per day is 5.
- deaths per day range from 0 to 5 (range = 5)

2. Try to describe in words the data presented in table 1.1

[illegible]

Summarizing data in this manner becomes difficult when the range in the data set is wide. In such situations the range has to be divided into groups or intervals.

3. Define a class interval

(Check your definition with what is given in this book)

Exercise 1.3

1. Let us look at the data set given in fig. 1.4. Does it differ from the earlier data set?

Heights(cm) of a sample of students in your batch

160.0	157.5	151.1	151.1
146.7	168.9	156.2	172.7
160.0	156.2	172.7	157.5
154.9	147.3	154.9	154.9
168.9	152.4	158.8	155.6
152.4	160.0	144.8	160.0
161.9	154.9	172.1	152.4
157.5	157.5	159.4	159.4
157.5	153.7	158.0	

Figure 1.4

1.3 Steps in summarizing data into a grouped frequency distribution

1. Find the range in your data set.

   min 144.8 .... max 172.7 .... range 27.9 ....

2. Divide the range into segments called " classes " or " class intervals ".

☞ As there are only 35 observations in your data set you don't need to have too many classes.

Write your class intervals here.

✍ 140 – 144.9
145 – 149.9
150 – 154.9
155 – 159.9
160 – 164.9
165 – 169.9
170 – 174.9

☞ number of classes could vary between 5 - 12 depending on the size of the data set

- minimum number of classes should be 5
- class limits in class intervals should not overlap
- classes are generally of equal size

Purpose of summarizing data is lost if there are many class intervals While, too few may make you lose important information about the data

3. Divide the range by the number of classes you choose to have and the width of a class can be determined.
4. List the class intervals in order and write them in column 1 of figure 5.1.
5. Take each observation from your data set and allocate it to the respective class it belongs to using a tally mark and fill column 2 of figure 5.1.

☞ Make sure that you have included all values. Check whether both "mums" (minimum and maximum) have been included into the first and the last class respectively.

☞ By re-arranging the raw data you are now able to describe the data using minimum and maximum values and the range.

2. What are the uses and limitations of re-arranging data into an array ?

✍ ....

3. Is it practical to re-arrange a large set of data in this manner ?

✍ ....

4. Is there a better way of organizing the data ?

✍ ....

Yes ! the data could be organized into a frequency distribution.

1.2. Steps in organizing data into a frequency distribution.

1. Arrange all possible values this variable could take (number of deaths per day) in an ascending order of magnitude

   If you examine column 1 in figure 1.3 you would realise that this column includes all possible values that fall within the range

2. Allocate each observation in the data set using a tally mark, to the respective category of the variable they belong to as shown in column 2

3. Count the tally marks in each category to obtain the frequency in column 3

column 1	column 2	column 3
Figure 1.3		
Number of Deaths/day (x)	Tally mark	Frequency (f)
0	~~////~~	5
1	~~////~~ //	7
2	~~////~~ ///	8
3	~~////~~ //	7
4	//	2
5	/	1

☞ The data has now being organized into a **Frequency Distribution**. This provi a summary of the raw data. This can be presented in a frequency distribu table as in table 1.1

Table 1.1

Distribution of number of deaths that occurred per day during a period of one month in hospital A

Number of deaths / day	Frequency (percent)
0	5 (16.7)
1	7 (23.3)
2	8 (26.7)
3	7 (23.3)
4	2 (6.7)
5	1 (3.3)
Total	30 (100)

Exercise 1.2

1. What are the uses and limitations of summarising data ?

✍ ....

....

....

....

....

6. Count the tally marks in each class interval and write the class frequencies in column 3.

☞ In a grouped frequency distribution the categories of the variable are in groups as opposed to what we saw in table 1.1.

column 1 column 2 column 3

Figure 1.5

Class intervals	Tally marks	Frequency

7. Present the heights you summarized above in the form of a table.

✍ Table 1.2

8. List the advantage of grouping data in a frequency distribution?

1.4 A table should have the following characteristics:

1. **A title,** clear and to the point.
2. **A number.**
3. **Headings** for each row and each column.
4. **Units of measurement** specified either in the title or row / column heading, as appropriate.
5. Totals for both the frequencies and percentage frequencies.

☞ A table has to be self-explanatory

Check the table 1.2 for the above features.

1.5 You have now learnt to **organize** and **condense** data.

The data has now:

- become **concise** without loosing the details
- brought out its pattern of variation
- become **simple and meaningful** enough to form impressions
- become **helpful** in further analysis
- been presented in a form that needs few words to explain

Data can be presented in many other ways that **arouses** the **interest** of the reader.

1.6 Methods of data presentation

You have already presented the data on the number of deaths in a hospital and heights of students in your batch in the form of tables.

The same data could also be presented by a **diagram**

1.6.1 Tabulation

In a frequency distribution table the variable is presented with its frequency and its percentage frequency.

Table 1.3

Distribution of number of deaths that occurred per day during a period of one month in hospital A

Number of deaths/day	Frequency (percent)
0	5 (16.7)
1	7 (23.3)
2	8 (26.7)
3	7 (23.3)
4	2 (6.7)
5	1 (3.3)
Total	30 (100)

By examining the frequencies and percentage frequencies the pattern of deaths in hospital A could be described.

☞ Percentage frequencies also become important when a comparison has to be made on occurrence of deaths between two hospitals.

By looking at table 1.3 the frequency of deaths in hospital A. could be described as follows:

- in eight out of 30 days i.e. 26.7% of the time there had been two deaths per day in hospital A. As this is the highest frequency seen in this data set we can say that, majority of the time there had been 2 deaths per day
- 3 deaths and 1 death per day had occurred with equal frequency i.e. 23.3% of the time
- sixteen point seven percent (16.7%) of the time, on 5 days, there has been no deaths

- four and 5 deaths per day have occurred 6.7% and 3.3% of ... respectively

Look at your table 1.2 on the frequency distribution of heights and try to des... distribution of heights

....
....
....
....
....
....
....
....

Exercise 1.4

Table 1.4 gives the frequency distribution of deaths seen in hospital B.

Table 1.4

Distribution of number of deaths that occurred per day during a period of one month in hospital B

Number of deaths/day	Frequency (percent)
0	5 (16.7)
1	9 (30.0)
2	10 (33.3)
3	5 (16.7)
5	1 (3.3)
Total	30 (100)

Compare the frequency of deaths in hospital A & B.

....
....
....
....
....
....
....
....
....

Table 1.5 shows the tabulation of heights of students. This table has certain additional features:

* cumulative frequency
* relative cumulative frequency

By adding up the frequency of the first class interval to that of the subsequent class interval, you get the **cumulative frequency** for the second class interval. This, when repeated gives the cumulative frequency for each class interval as given in column 3 of table 1.5. Similarly, the **relative cumulative frequencies** for class intervals are obtained by cumulating the relative frequencies.

Table 1.5
Frequency, cumulative frequency and relative cumulative frequency distributions of a group of medical students in your batch.

column 1 Heights (cm)	column 2 Frequency (percent)	column 3 Cumulative frequency	column 4 Relative cumulative frequency
140.0-144.9	01 (2.9)	01	2.9
145.0-149.9	02 (5.7)	03	8.6
150.0-154.9	09 (25.7)	12	34.3
155.0-159.9	13 (37.1)	25	71.4
160.0-164:9	05 (14.3)	30	85.7
165.0-169.9	02 (5.7)	32	91.4
170.0-174.9	03 (8.6)	35	100.0
Total	35 (100)		

Well !, I can describe the distribution of heights presented in table 1.5 as.

- 37 % of the students (which is the highest frequency) have their heights between 155.0 and 159.9 cm
- followed by 25.7% of students whose heights are distributed between 150.0 to 154.9 cm .... and so on
- looking at the relative cumulative frequency, 71.4% of the students have their heights equal to or less than 159.9 cm and so on ....

1.6.2 DIAGRAMS

Data presented in a table could be presented graphically or diagrammatically. When graphing data, it is customary to indicate the values of the variable on the horizontal or x axis and the frequency of occurrence on the vertical or y axis.

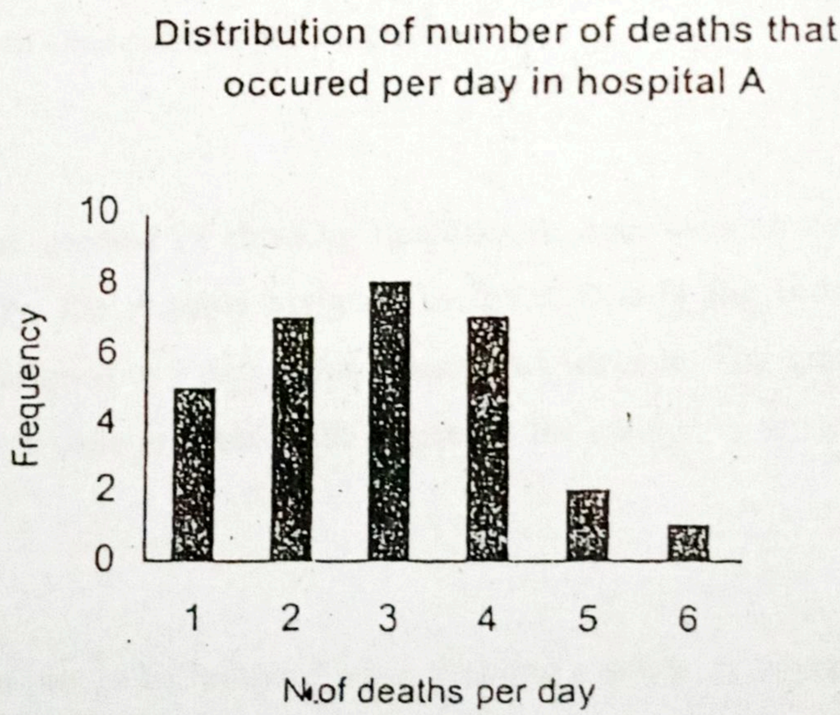
There are many forms of diagrams. Some of them are:

1. Bar diagram
2. Histogram
3. Frequency polygon
4. Frequency curve
5. Pie chart
6. Graph

Let us now look at how we could present data in table 1.1. diagrammatically.

BAR DIAGRAM

Figure 1.6



* The variation in the occurrence of deaths per day is very clearly visualised here.

1.6.2.1 Steps in drawing a bar diagram

1. The length of bars represent the frequency or percentage frequency. Each bar is drawn proportional to the frequency or percentage frequency
2. The width of bars should be uniform throughout the diagram

☞ There are spaces between bars.

This is a **simple bar diagram.**

Multiple and composite bar diagrams too could be drawn.

Refer to the fixed learning module (FiLM) for multiple composite bar diagrams.

Look at your table on the frequency distribution of heights in table 1.2 and present this data diagrammatically:

☞ Hope you realise that the number of deaths per day, is a discrete variable and heights, is a continuous variable. A discrete variable may be presented by a Bar diagram while a continuous variable could be presented by a **Histogram.**

1.6.2.2 Steps in drawing a histogram

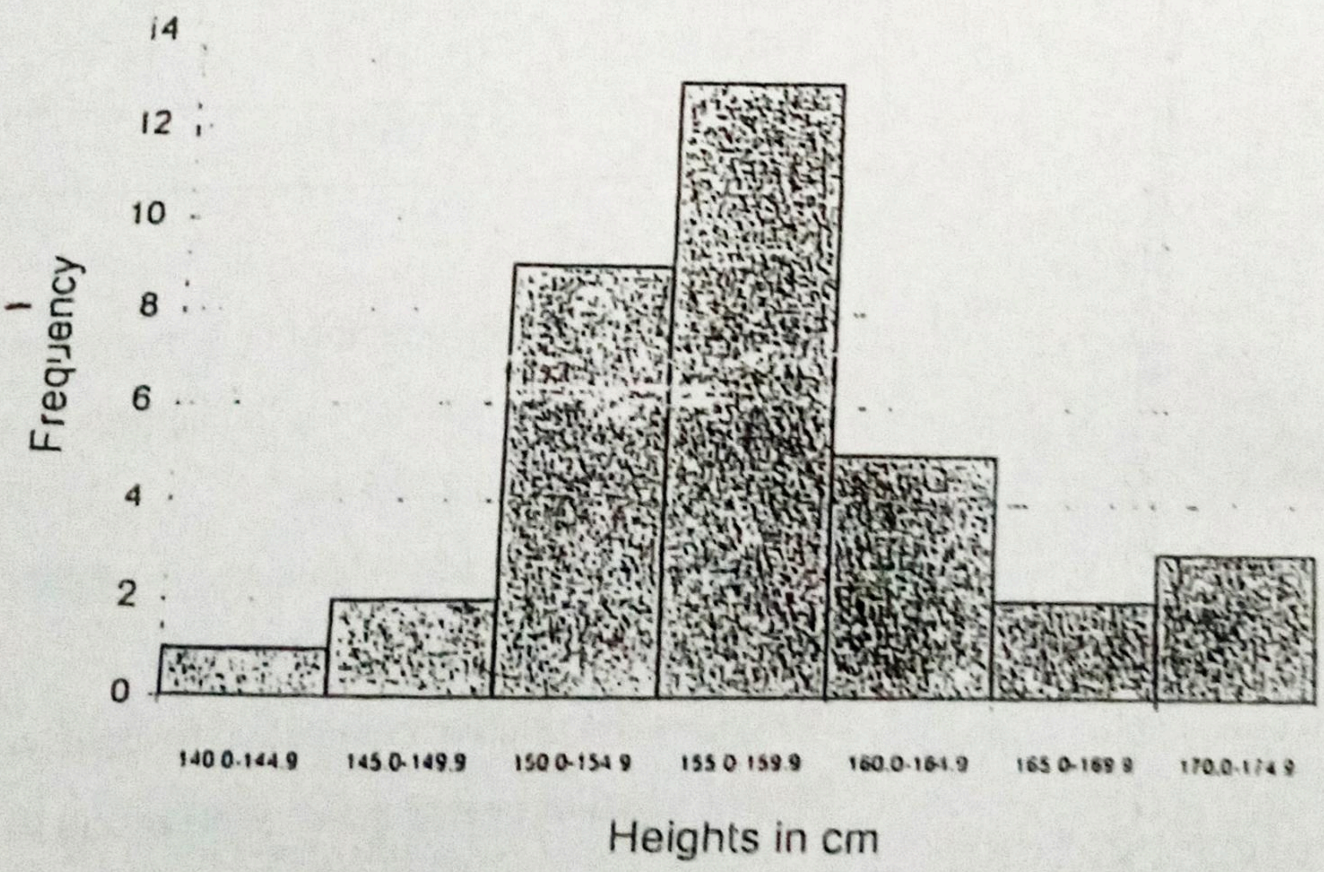
1. The horizontal axis, x, presents the variable and the vertical axis, y, the frequencies or the percentage frequencies.
2. The values of the variable are presented by drawing vertical bars proportional to frequencies or percentage frequencies. Here the vertical bars are pasted against one another. Marking and labeling of the horizontal axis could be done using the lower class limits, class mid-points or class boundaries.

☞ A convenient way to draw a histogram would be to draw it on class boundaries. As the class boundaries overlap automatically the rectangles get adjacent to one another.

Figure 1.7 presents the histogram of the frequency distribution of heights given in table 1.5.

Figure 1.7

Distribution of heights of a group of medical students



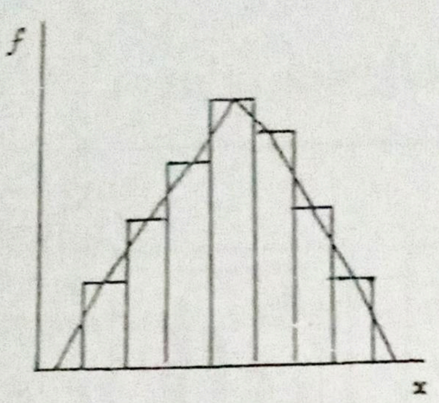


Figure 1.8

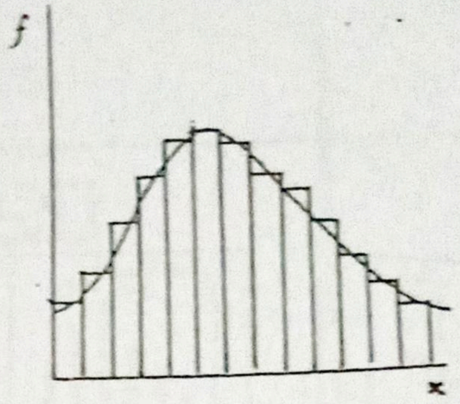


Figure 1.9

1.6.2.3 If you join the mid points of the upper horizontal lines of each of the rectangles in the histogram by a straight line you would get a **frequency polygon** as shown in figure 1.8.

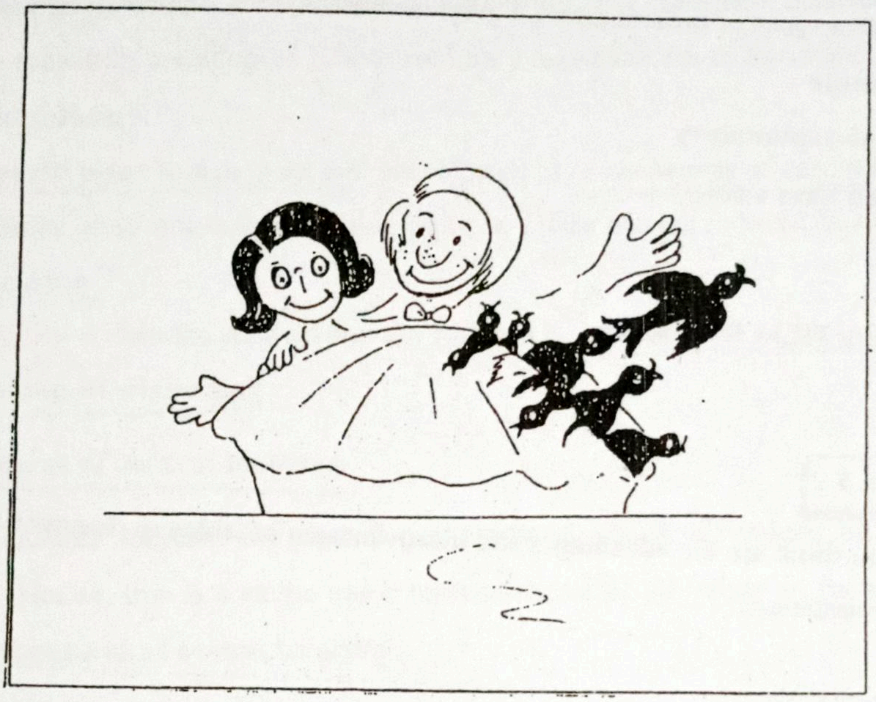
In a large data set when you have many class intervals with narrow width the frequency polygon takes the form a smooth curve as in figure 1.9. This is called a **frequency curve.**

☞ Frequency distribution of a continuous variable is diagrammatically present by a **histogram**, while a discrete variable is presented by a bar diagram.

We have looked at some forms of diagrammatic data presentations. There are many more.

1.6.2.4 **Pie diagram** or pie chart is another form of pictoral presentation of data.

☞ Like cutting up a pie into pieces, in a pie diagram the frequencies or the percentage frequencies of a variable is presented by dividing 360^0 of a circle, proportionately.



Refer FiLM, for an example of a pie-diagrams.

1.6.2.5 Graph

A graph is a diagrammatic method of showing quantitative data using a co-ordinate system, x and y. Generally, the variable assigned to the x axis is the **independent** variable and the variable assigned to y axis is the **dependent** variable. The graph shows the change in the value of variable in y axis with respect to the change in the variable in the x axis.

Certain basic rules should be followed when drawing a graph to avoid creating a misleading visual impressions.

Following are some important points to observe when drawing a graph.

1. It is important to draw a graph to scale, and to start the vertical axis from zero.
2. It is traditional to draw the vertical axis 75% in length of the horizontal axis.

Like for all other types of presentations, a graph should be,

- **simple**
- **self-explanatory**
- and have a title

Refer FiLM for graphs.

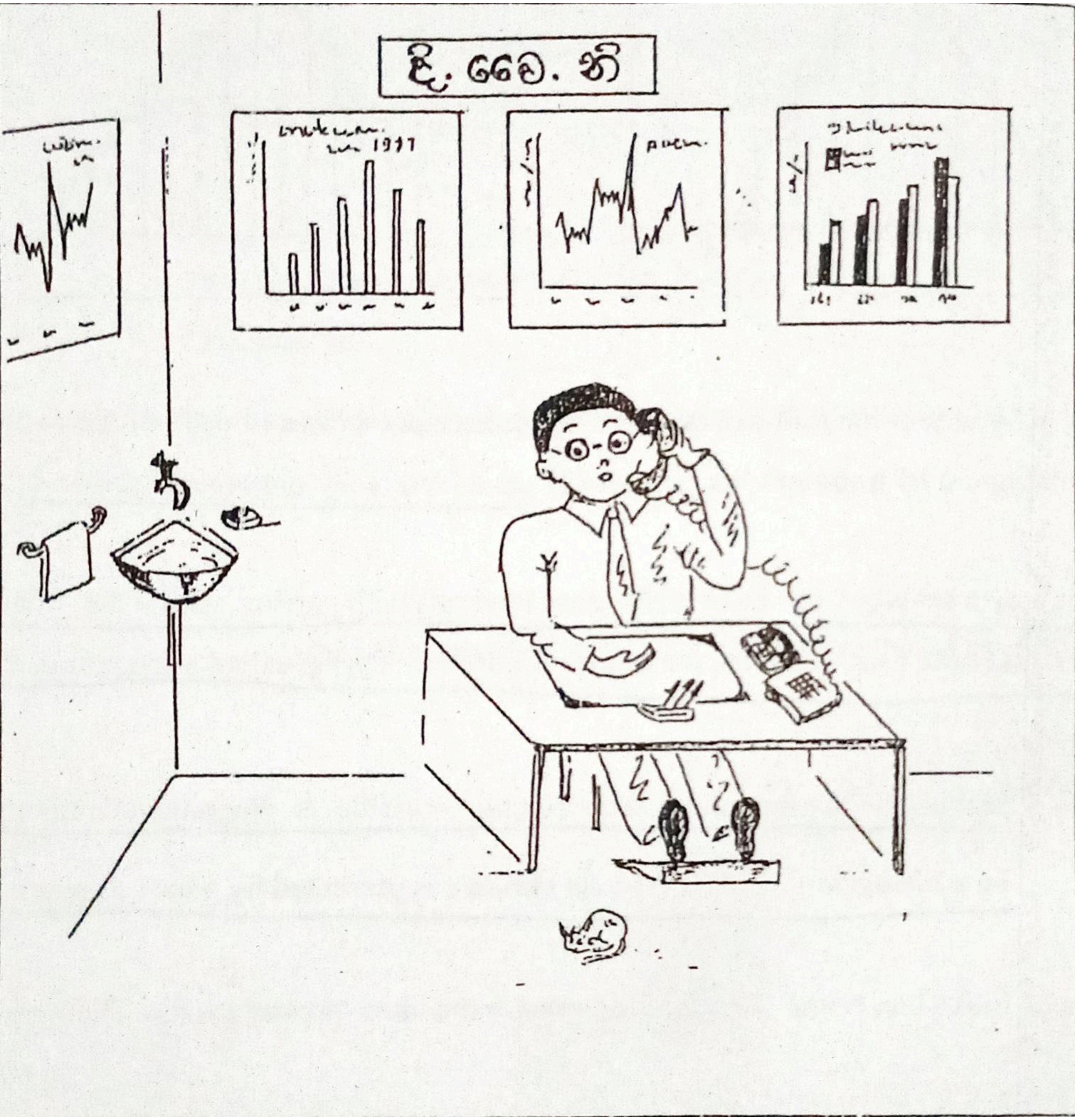
Exercise 1.5

What do you think are the advantages and disadvantages of tables vs diagrams as method of data presentation?

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

Both tables and diagrams have their own advantages and disadvantages as methods of presenting data.

* Tables are more scientific as they are precise.
* Diagrams creates a better visual impact.


ඊ. චෙ. ති
1977

I - DESCRIPTIVE STATISTICS

LESSON 2

Numerical methods of describing data.

LEARNING OBJECTIVES

At the end of the lesson the student should be able to:

1. Explain the following **summary measures** and their application in medical science.
   1.1 **measures of central tendency (averages) and measures of location.**
   1.2 **measures of dispersion.**
2. Calculate the above measures for a given set of data.
3. Select an appropriate measure of central tendency to describe a given set of data.
4. Explain the use of **centiles.**
5. Comment on the uses and limitations of such measures and their advantages and disadvantages.

2. Numerical measures

Biological measurements are subjected to variability. To describe a data set, appropriate summary measures are adopted. These summary measures are of two types.

Those that measure:

1. the central point in a data set and are referred to as **measures of central tendency** and those that measure any other point in a data set are referred to as **measures of location**
2. variability within the data set and are referred to as **measures of variability, variation or dispersion**

2.1 Measures of central tendency

These measures indicate the central point around which all values of a variable are arranged. Hence, this is a single value representative of all values in the data set. There are three measures of central tendency.

They are:

2.1.1 Mean

2.1.2 Median

2.1.3 Mode

2.1.1 Mean

There are different types of means of which the arithmetic mean is the most popular and widely used. There are two types of arithmetic means:

i. simple arithmetic mean

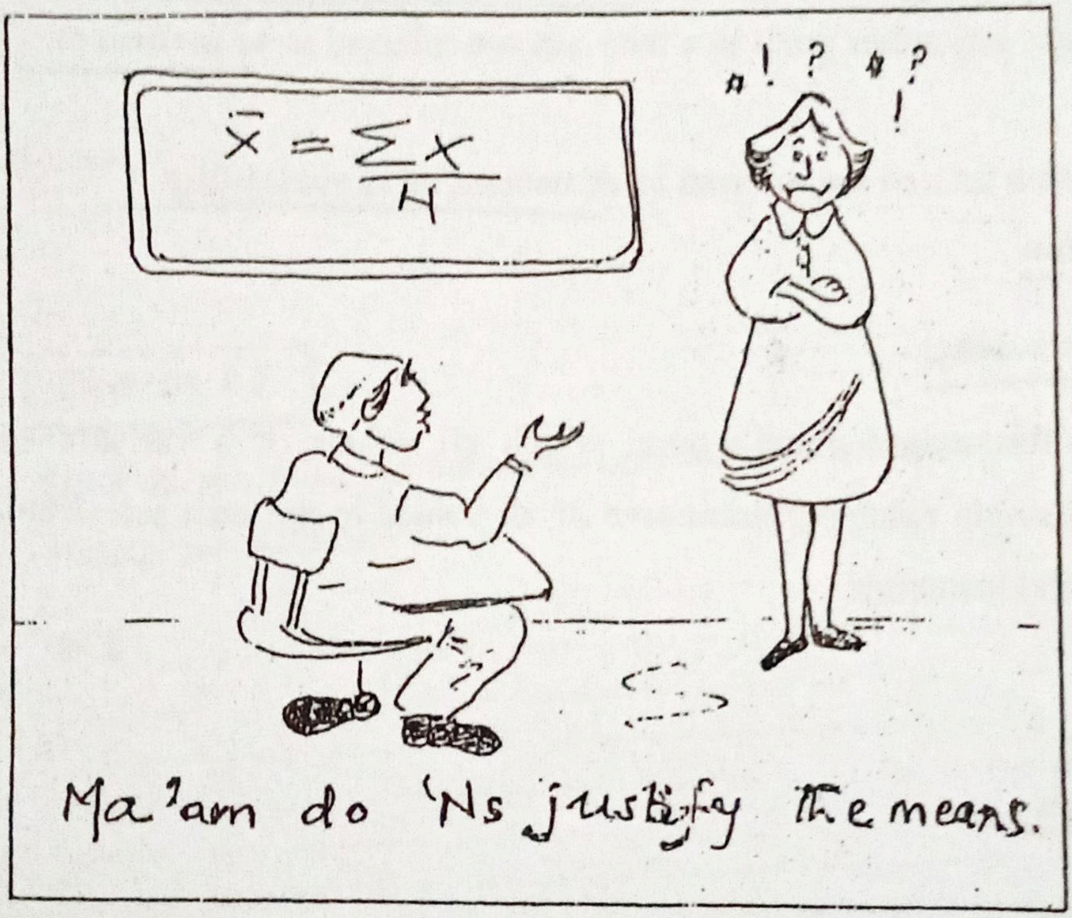
ii. weighted arithmetic mean

i. Simple arithmetic mean ($\bar{x}$)

This is the sum of all values of a variable divided by total number of values observations.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots\ldots\ldots\ldots x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \textit{formula 2.1}$$

x_1 to x_n are the values taken by the variable x. Hence, x_i refers to all values of the variab x. Sigma (Σ) is the operational term which indicates adding up of all values of x. The total number of values for variable x is n (number of observations).



Exercise 2.1

Marks scored at a continuous assessment examination by a group of 10 randomly selected students are:46 54 28 39 49 63 57 68 52 00

Exercise 2.2

Number of drugs given to patients attending an OPD is given in table 2.1

Table 2.1

Distribution of number of drugs given per patient in an OPD

Number of drugs (x)	Frequency (f)
0	8
1	42
2	8
3	2
4	0
Total	60

Using the above data calculate the average number of drugs given to a patient

$\bar{x} = \frac{42+16+6}{60} = \frac{64}{60} = 1.07$

If your answer is correct the method you used can be expressed as follows:

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{\sum fx}{n}$$

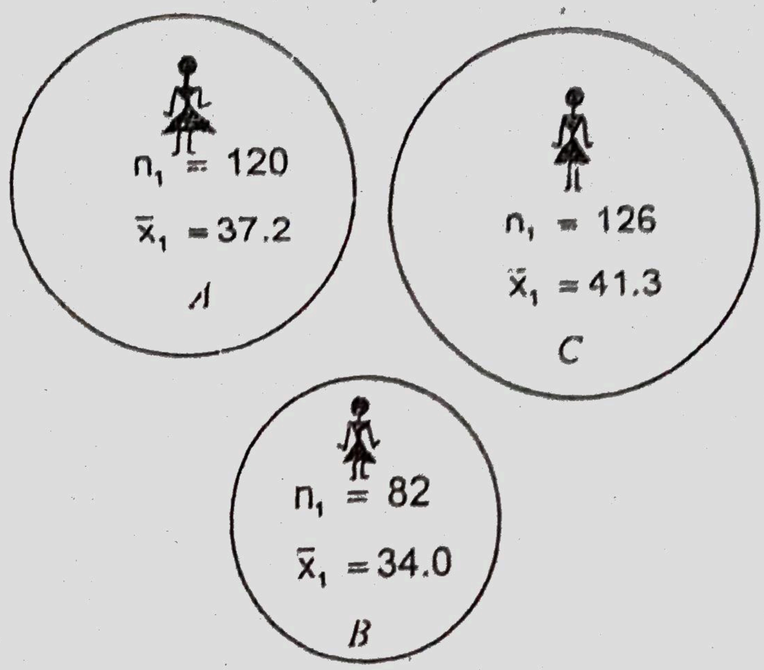
formula 2.2

Weighted arithmetic mean

Example 2.1

Figure 2.1 shows three groups of children. The number and the mean weight of children for each group is given.

Figure 2.1



Can we say that the mean weight of all the children are: $\dfrac{37.2 + 34.0 + 41.3}{3}$

$= 37.5$ kg ?

No, here the weights given are in respect of the three groups of children.

Total weight for each group of children and, all children in the 3 groups can be calculated as,

Total weight in group A = $37.2 \times 120 = 4464$

Total weight in group B = $34.0 \times 82 = 2788$

Total weight in group C = $41.3 \times 126 = 5204$

Total weight of all children = $4464 + 2788 + 5204 = 12456$

Number of children in groups A, B, and C = $120 + 82 + 126 = 328$

If we divide the total weight by the total number of children, we get the mean weight for the whole group as 37.98 kg.

This can also be represented as,

$$\bar{x}_k = \frac{w_1\bar{x}_1 + w_2\bar{x}_2 + \ldots\ldots\ldots w_k\bar{x}_k}{w_1 + w_2 + \ldots\ldots\ldots w_k} = \frac{\sum_{i=1}^{k} w_i\bar{x}_i}{\sum_{i=1}^{k} w_i} \qquad \textit{formula} \quad 2.3$$

A set of n observations consist of k different groups where x_i and w_i are the average and frequency for each group (i =1,2,......k). In calculating the mean for the entire group, average for each group is multiplied (weighted) by the number of observations (frequency) in each group.

How is the mean calculated on data given in the form of a grouped frequency distribution? Let us look at the example given below.

Exercise 2.3

The birth weights of 470 infants born in hospital A during one year is given in table 2.2. Find the mean birth weight of the infants.

Table 2.2

Distribution of infants by their birth weights

Birth weight of infants (kg)	No. of infants (f)
2.0 - 2.4	15
2.5 - 2.9	96
3.0 - 3.4	190
3.5 - 3.9	135
4.0 - 4.4	28
4.5 +	6
Total	470

When data are grouped into class intervals the individual weights of 470 infants or the raw data are not available. We assume that the average birth weight for each class is the mid-point of that class interval. The class mid point multiplied by the frequency of observations in that class will give a value which would approximate the sum of the original birth weights of infants falling into that interval.

2.1.1a Steps in calculating the mean from grouped data

1. Find the class mid-point for each class interval. The class mid-point is considered as the average value ($\bar{x}$), for that class interval.
2. Multiply the class mid-point by its frequency (f). This would give the estimated total for the group. Complete the columns given in the table 2.3.

column 1 column 2 column 3 column 4

Table 2.3

Distribution of infants by their birth weight

Birth weight of infants (kg)	No. of infants (f)	Class mid point (x)	(fx)
2.0 - 2.4	15	2.2	33
2.5 - 2.9	96	2.7	259.2
3.0 - 3.4	190	3.2	608
3.5 - 3.9	135	3.7	499.5
4.0 - 4.4	28	4.2	117.6
4.5 +	6	4.7	28.2
Total	470		Σfx 1545.5

$\bar{x} = \frac{[illegible]}{[illegible]}$

= 3.[illegible]

3. Sum up column 4 (Σfx).
4. Divide Σfx by the total number of observations (n). This gives the mean weight of infants born in hospital A.

☞ This method is a short cut method to calculate the mean when the data set is large. This method has been used to calculate the mean before calculators came into use.

Now you know how to calculate the mean when data are given in a series of values and when summarised into a frequency distribution (grouped and ungrouped).

2.1.2 Median

The **median** is the value which occupies the middle position when all observations are arranged in an array, either ascending or descending.

Exercise 2.4

Erythrocyte sedimentation rates in (mm) of 7 subjects are given below:

7, 5, 3, 4, 6, 4, 5

✍ 7, 6, 5, 5, 4, 4, 3

Median = 5 mm

What is the median?...............

2.1.2a Steps in calculating the median

1. Arrange the values in an array.
2. Number the observations in serial order as follows.

1st	2nd	3rd	4th	5th	6th	7th
3	4	4	5	5	6	7

☞ As there are only 7 observations, the middle observation can easily be located. However, when the data set is large, the expression, $(n + 1) / 2$ can be used to find the serial number of the median value.

3. Use the formula to find the median.

Exercise 2.5

Erythrocyte sedimentation rates of 8 subjects are given below:

7, 5, 3, 5, 4, 6, 4, 5

Calculate the median

7, 6, 5, 5, 5, 4, 4, 3.

Median = 5 mm

You would have realised that for an even number of observations, there are two middle numbers.

The equation when applied to this data gives the following 8+1/2 = 4.5. This indicates that the median is located between the 4th and 5th observations. The median is thus the average of the observations corresponding to these two serial numbers, i.e. (5+5)/2 = 5. The expression given below could be used to find the median in grouped data:

$$\text{Median} = L + \frac{(n/2 - F)\,C}{f} \qquad \textit{formula } 2.4$$

where, L, is the lower class boundary of the median class

n, the total number of observations

F, cumulative frequency of the class above the median class

f, frequency of the median class

C, width of the median class

Exercise 2.6

Let us look at how we could apply formula 2.4 to find the median of the data on birth weights of infants given in table 2.2.

2.1.2b Steps in calculating the median from grouped data

1. Identify the class interval to which the median belongs.

In order to find the median class, use the expression (n+l) / 2. (471/2 = 235.5), i.e. both 235[th] and 236[th] observations are middle numbers. Next is to see to which class these two numbers belong. In order to find that it is necessary to find the cumulative frequency. Fill column 3 in table 2.4. Looking at the cumulative frequencies you should realise that both 235th and 236th observations lie in the 3rd class interval, 3.0 - 3.4. This is the class to which the median belongs.

Table 2.4

Distribution of infants by their birth weight

column 1 Weight of infants (kg)	column 2 No. of infants (f)	column 3 Cumulative frequency (F)
2.0 - 2.4	15	15
2.5 - 2.9	96	111
3.0 - 3.4	190	301
3.5 - 3.9	135	436
4.0 - 4.4	28	464
4.5 +	6	470
Total	470	

2. Substitute values to formula 2.4 to find the median.

Median = 2.95 + $\frac{(470/2 - 111)}{190}$ × 0.5 = 2.95 + [illegible]

= 3.28

2.1.3 Mode

Mode is the most commonly occurring value in the data set or it is the value that occurs with the greatest frequency. In a frequency curve mode refers to the highest point.

(17)

Exercise 2.7

Table 2.5 shows the distribution of drugs given per patient in an Out Patient Department

What is the mode in the following distribution..

Table 2.5

Distribution of number of drugs given per patient in an OPD

Number of drugs (x)	Frequency (f)
0	8
1	42
2	8
3	2
4	0
Total	60

We have already learnt that an average is a single value representing a group of values. For a value to be a "good" average it should satisfy certain criteria. An average should :

* be easily understood
* be simple to compute
* be based on all values
* not be unduly affected by extreme values
* be defined by a mathematical formula
* be capable of further algebraic treatment

Taking each of the above properties into consideration, do you think all three averages are equally "good" in representing the data? Let us try to understand this by looking at some examples.

Example 2.2

Haemoglobin levels (g/dl) of 26 normal children are as follows:

11.8	10.8	10.8	12.4	11.6	13.2	13.8
11.4	12.2	12.0	11.7	12.6	14.2	12.2
10.4	12.9	10.5	12.7	13.3	13.5	11.6
12.3	11.2	12.2	12.9	13.0		

If you calculate the mean, median and mode for this data, you will realise that all three averages have the same value 12.2 g/dl. Hence, any one of them could be used as a value representing the data. However, referring to the properties of a " good" average, mean is easily understood and simple to compute. Neither arraying of data as required for calculating the median nor grouping of data as required for locating the mode are necessary to calculate the mean. Mean is based on all values in the data set including extreme values, whereas both the median and the mode refer to one value in the data set. Being defined by a mathematical formula it can be subjected to further algebraic treatment when compared with the median and the mode. However, the mean has a limitation when used in representing data.

What do you think is the limitation ?

....

....

....

Exercise 2.8

A health planner wants to open a new unit in hospital B and needs to know the average duration of stay by patients in this hospital in order to calculate the optimum number of beds necessary. He has studied the duration of stay of patients admitted during the past one month to this hospital and the data obtained is given in table 2.6

Table 2.6

Duration (in days)	1	2	3	4	5	6	7	8	9	10
No. of patients	20	77	30	21	15	11	9	7	7	5

1) Calculate the different averages,
2) Choose the value that you would use if you were the health planner.
3) Give reasons for your choice.

....
....
....
....
....
....
....

Exercise 2.9

Monthly income of 11 employees in different occupations working in a private sector organization is as follows:

Rs.12,000/=	Rs.9,000/=	Rs. 2,000/=	Rs.3,000/=
Rs.10,000/=	Rs.6,000/=	Rs.25,000/=	Rs.9,500/=
Rs. 2,000/=	Rs.3,000/=	Rs. 3,000/=	

The mean, median and mode when calculated works out to:
mean = Rs. 7,681.82, median = Rs.6,000/=, mode Rs.3,000/=. Which of the above three measures do you think is the best to describe the monthly income of employees? Give reasons for your choice.

....
....
....
....
....
....
....
....
....

2.2 Measures of location

Averages or measures of central tendency locate the center (mid point) of the distributi of a variable. It would provide more information if the data is also described givi values of the variable at other points. In order to locate these points it is necessary arrange the values in an ascending order of magnitude or draw a frequency curve. (Ref page 19).

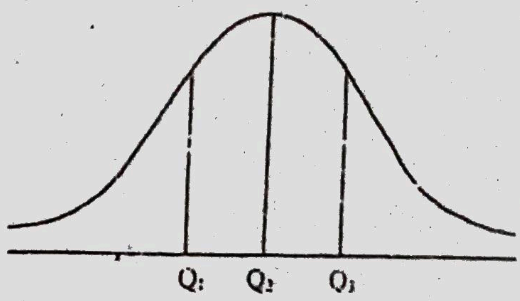
Commonly used measures of location are :

- i Quartiles
- ii Deciles
- iii Centiles or Percentiles

i. Quartiles

Quartiles divide a distribution into four equal parts. There are three quartiles.

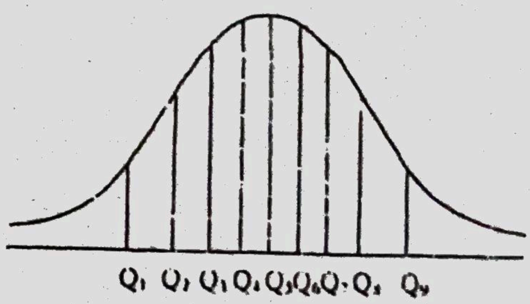
Figure 2.2



ii. Deciles

Deciles divide a distribution into 10 equal parts. There are nine deciles.

Figure 2.3



iii. Centiles or percentiles

Centiles or percentiles divide the distribution into 100 equal parts. The median is the 50th percentile. Percentiles are good measures of locations. They are used to describe the growth in children.

2.3 Measures of Dispersion

Variability is a normal characteristic of most biological phenomena. The mean though influenced by the range of observations, does not tell us anything about the degree of dispersion observed in the raw data. Therefore, we need to compute other measures to describe the variability observed.

Example 2.3

Haemoglobin levels of two groups of children are:

Group I: 12.1, 12.2, 12.8, 12.9, 12.3, 12.4, 12.7, 12.6, 12.5

Group II: 12.1, 12.3, 11.7, 11.9, 13.1, 13.3, 12.5, 12.9, 12.7

The range and average for the above data is as follows:

	Group I	Group II
No. of observations	9	9
Average	12.5	12.5
Range	0.8	1.6

This tells you that the average alone is not good enough to describe data. The dispersion or scatter of the data too, has to be looked at when describing data.

Measures of Dispersion are:

2.3.1 Range
2.3.2 Inter-quartile range
2.3.3 Mean deviation
2.3.4 Standard deviation
2.3.5 variance
2.3.6 Co-efficient of variation

2.3.1 Range

You already know what range is. It is the difference between the largest and the smallest values in the data set. Look at the haemoglobin levels of children in Group I and Group II in example 2.3. The range could be given as follows:

Group I - 12.1 to 12.9 and / or 0.8

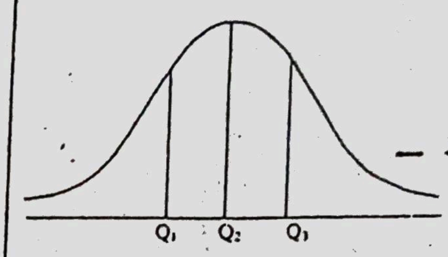
Group II - 11.7 to 13.3 and / or 1.6

Range is the simplest measure of dispersion. It only tells the minimum and maximum values and that all other values are placed between them.

2.3.2 Inter-quartile range

You already know what a quartile is. The inter-quartilerange is the range within which the middle 50% of thevalues lie.

Figure 2.4



2.3.3 Mean deviation

The mean deviation is a better measure than the above two. It is the arithmetic average of the deviation of observations from the arithmetic mean. The formula to calculate the mean deviation could be given as:

$$\text{Mean deviation} = \frac{\Sigma(x-\bar{x})}{n} \qquad \textit{formula } 2.5$$

As observations are distributed both above and below the mean, the differences of each observation from the mean may be either a positive or a negative value, the sum of these differences equaling zero. In calculating the mean deviation, all differences are considered as positive.

Steps in calculating mean deviation :

1. Calculate the mean for the data
2. Find the difference of each observation from the mean (ignore the sign, take all differences as plus)
3. Add the differences.
4. Divide by the number of observations

Though the mean deviation is a better measure of dispersion than the range or the inter-quartile range it is not used in statistical analysis in drawing inferences.

(21)

Example 2.4

The Erythrocyte Sedimentation Rate (ESR in mm) of 8 normal individuals are: 4, 4, 5, 4, 3, 4, 5, 3Find the mean deviation.1. Enter the data into column 1 as shown in table 2.7

column 1 column 2

Table 2.7

Erythrocyte sedimentation rate and the deviation of each observation from the mean

ESR (x)	Deviation from mean $(x-\bar{x})$
4	0
4	0
5	1
4	0
3	1
4	0
5	1
3	1
Total 32	4

2. Find the mean.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{32}{8} = 4$$

3. Subtract the mean from each value as shown in column 2.
4. Add all differences.

$$\sum(x-\bar{x}) = 4$$

5. Divide the answer you observed in step 4 by the total number of observations.

$$\frac{\Sigma(x-\bar{x})}{n} = \frac{4}{8} = 0.05$$

$$\text{mean deviation} = 0.05$$

2.3.4 Variance and

2.3.5 Standard deviation (SD)

Variance and standard deviation are better measures than the mean deviation and a most commonly in statistical analysis. Formula to calculate the variance and st deviation could be given as:

$$\text{Variance} = \frac{\Sigma(x-\bar{x})^2}{n-1} \qquad \textit{formula 2.6}$$

$$\text{Standard deviation} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} \qquad \textit{formula 2.7}$$

Steps in calculating the standard deviation and variance

1. Enter the observations vertically downwards in a column.
2. Find the mean.
3. Find the difference of each observation from the mean and enter alongside each observation in a second column.

☞ when calculating the standard deviation and variance the sign of the devi has to be taken into account.

4. Square the differences of each observation from the mean and enter in a column alongside with the differences.
5. Add the squared values to get the sum of squares.
6. Divide this sum by the number of observations minus one to get the mean sq deviation. This is called the **variance.**
7. Find the square root of this variance to get the **standard deviation.**

Variance and Standard deviation can be expressed by another formula:

$$\text{Variance} = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} \qquad \textit{formula 28}$$

$$\text{Standard deviation} = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}} \qquad \textit{formula 29}$$

☞ A large standard deviation indicates that the values of the variable are widely spread out from the mean. A small standard deviation means that the values are scattered closely around the mean.

Exercise 2.10

Respiratory rate in 9 patients are:

23, 22, 20, 24, 16, 17, 18, 19, 21

Find the standard deviation and variance of the respiratory rates. Use table 2.8

Table 2.8

Respiratory rate, deviation of each value from the mean and the squared deviation

Respiratory rate (res. / min.) x	Deviation from mean $(x - \bar{x})$	Squared deviation $(x - \bar{x})^2$
23	+ 3	9
22	+2	4
20	0	0
24	+4	16
16	−4	+16
17	−3	+ 9
18	−2	+ 4
19	−1	1
21	+ 1	1
180	0	60
Total $\bar{x} = 20$		

7.5

23

....

23

You must also learn to calculate the standard deviation and variance in grouped data. Try the exercise given below.

Exercise 2.11

Intelligence quotient (IQ) of 50 boys is given below. Find the standard deviation and variance. Use both formulae.

Table 2.9

IQ	f	Class midpoint x	fx	Deviation from $\bar{x}$ $x-\bar{x}$	Squared deviation $(x-\bar{x})^2$	$f(x-\bar{x})^2$
0 - 19	3					
20 - 39	4					
40 - 59	3					
60 - 79	4					
80 - 99	13					
100 -119	12					
120 -139	8					
140 -159	3					
Total	50					

When data is given in the form of a frequency distribution, rememb... value taken by the variable in each class interval (mid-point) has to... multiplied by the frequency. Each deviation too has to be multipli... frequency.

$$\text{Variance} = \frac{\Sigma f(x-\bar{x})^2}{n-1} \qquad \textit{formula 2.}$$

$$\text{Standard deviation} = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n-1}} \qquad \textit{formula 2.1}$$

$$\text{Variance} = \frac{\Sigma x^2 f - \frac{(\Sigma fx)^2}{n}}{n-1} \qquad \textit{formula 2.1}$$

$$\text{Standard deviation} = \sqrt{\frac{\Sigma x^2 f - \frac{(\Sigma fx)^2}{n}}{n-1}} \qquad \textit{formula 2.1}$$

8.3.6 Coefficient of variation (CV)

It is a measure used to compare relative variability. It may be the variation of... characteristic in two or more different series, e.g. variation of pulse rate among ... and old. It can also be used to measure two different characteristics in the sa... e.g. heights and weights among medical undergraduates or pulse and blood press...

Co-efficient of variation is a measure that compares variability irrespective of the unit of measurement, e.g. pulse rate and blood pressure (mm of Hg).

$$CV = \frac{SD}{mean} \times 100$$

→ P. 45

DEFINITIONS

Class interval – an interval of values into which a range of values of a variable has been divided

Class limit – the end numbers of a class interval. A class interval has two class limits: Lower class limit and upper class limit

Class boundaries – true class limits of a class interval

Class width – the distance from the lower class boundary to the upper class boundary in a class interval

Percentage frequency – frequency of a class when presented as a percentage of the total frequency

Independent and dependent variables – When an association between two variables is studied, the variables may be referred to as the independent and dependent variables. The variable that is used to describe or measure the problem under study is called the dependent variable. The variables that are used to describe or measure the factors that are assumed to cause or influence the problem are called the independent variable. In a study of the relationship between smoking and lung cancer, "suffering from lung cancer" would be the dependent variable and "smoking" the independent variable.

ANSWERS TO EXERCISES

2.1. 45.6
2.2. 1.07 drugs
2.3. 3.29 kg
2.4. 5 mm
2.5. 5 mm
2.6. $2.95 + \frac{(470/2-111)}{190} \times .5 = 3.28\text{kg}$

2.7. 1 drug
2.8. i) mean= 3.57 days, median = 3 days, mode = 2 days
ii) 2 days (mode)
2.10. variance = 7.5, standard deviation = 2.74
2.11 variance = 1394.45, standard deviation = 37.34

CV is expressed as a percentage and is given by:

$$CV = \frac{SD}{Mean} \times 100$$ *formula* 2.14

Example 2.5

In a group of boys the mean systolic blood pressure in mm Hg. was 120 and SD was 10. In the same group of boys the mean height and SD was 160 cm and 5 cm, respectively. Which characteristic shows greater variation?

$$\text{CV of blood pressure} = \frac{10}{120} \times 100 = 8.3\%$$

$$\text{C V of heights} = \frac{5}{160} \times 100 = 3.1\%$$

Blood pressure is found to show a greater variability i.e. 2.7 times that of the height (8.3/3.1).

Annex 3

Proof of the formula for the binomial variance

Suppose a population of N individuals is classified as 'success' or failure, an each individual define x=1 for a 'success' and 0 for a failure. The mean value μ is, in fact, the population proportion of success, π.

$$\text{Thus } \mu = \frac{\Sigma x}{N} = \frac{\text{Popn. number of successes}}{N} = \pi$$

And the population variance of x is $\Sigma (x - \mu)^2 / N$

$$= \frac{\Sigma x^2 - (\Sigma x)^2/N}{N} = \frac{\Sigma x^2}{N} - \frac{(\Sigma x)^2}{N^2} = \pi - \pi^2 = \pi(1 - \pi)$$

Suppose a sample of N individuals is taken from the population N individuals. number of successes out of n say r is simply the sum of the n values of x.

So. $var(r) = var(\Sigma x) = \Sigma \; var(x) = \Sigma \pi(1 - \pi)$

Since the summation is over all n sample values

$$Var(r) = n\pi(1 - \pi)$$

And thus the estimated variance of r is obtained by replacing the population prop π by the sample proportion p. Thus the estimated variance of r is $np(1-p)$ whic sometimes written as npq where $q=1-p$ is the proportion of failure.

ANSWERES TO EXERCISES

3.1. 1. 4/52 2. 1/6

3.2. 1. 80/200 2. 120/200

3.3. 34 Students

3.4. 32 Soldiers

3.5. 1. 15.86 % 2. 0.13 %

4.4. 127.72 and 129.88

4.5. Boys, 80.95 and 101.45
Girls, 82.16 and 99.44

4.7.1 0.46, 95% C.I. = 0.32 to 0.59
99% C.I. = 0.28 to 0.64

5.1. No, SND = 10, $P<0.01$

5.2. No, SND = 2.38, $P<0.05$

5.3. Yes, SND = 1, $P>0.05$

5.4. 1. 0.42 2. Yes 3. 0.001

5.5. Yes, SND = 6.0, $P<0.01$

5.6. Yes, SND = 7.59, $P<0.01$

5.7 Yes, SND = 2, $P<0.05$

Annex I

The Standard Normal Distribution

SND		0	1	2	3	4	5	6	7	8	9
0.0	0.	50000	49601	49202	48803	48405	48006	47608	47210	46812	46414
0.1		46017	45620	45224	44828	44433	44038	43644	43251	42858	42465
0.2		42074	41683	41294	40905	40517	40129	39743	39358	38974	38591
0.3		38209	37828	37448	37070	36693	36317	35942	35569	35197	34827
0.4		34458	34090	33724	33360	32997	32636	32276	31918	31561	31207
0.5		30854	30503	30153	29806	29460	29116	28774	28434	28096	27760
0.6		27425	27093	26763	26435	26109	25785	25463	25143	24825	24510
0.7		24196	23885	23576	23270	22965	22663	22363	22065	21770	21476
0.8		21186	20897	20611	20327	20045	19766	19489	19215	18943	18673
0.9		18406	18141	17879	17619	17361	17106	16853	16602	16354	16109
1.0		15866	15625	15386	15151	14917	14686	14457	14231	14007	13786
1.1		13567	13350	13136	12924	12714	12507	12302	12100	11900	11702
1.2		11507	11314	11123	10935	10749	10565	10383	10204	10027	*98525
1.3	0.0	96800	95098	93418	91759	90123	88508	86915	85343	83793	82264
1.4		80757	79270	77804	76359	74934	73529	72145	70781	69437	68112
1.5		66807	65522	64255	63008	61780	60571	59380	58208	57053	55917
1.6		54799	53699	52616	51551	50503	49471	48457	47460	46479	45514
1.7		44565	43633	42716	41815	40930	40059	39204	38364	37538	36727
1.8		35930	35148	34380	33625	32884	32157	31443	30742	30054	29379
1.9		28717	28067	27429	26803	26190	25588	24998	24419	23852	23295
2.0		22750	22216	21692	21178	20675	20182	19699	19226	18763	18309
2.1		17864	17429	17003	16586	16177	15778	15386	15003	14629	14262
2.2		13903	13553	13209	12874	12545	12224	11911	11604	11304	1101[illegible]
2.3		10724	10444	10170	*99031	96419	93867	91375	88940	86563	842[illegible]2
2.4	0.0^2	81975	79763	77603	75494	73436	71428	69469	67557	65691	63[illegible]72
2.5		62097	60366	58677	57031	55426	53861	52336	50849	49400	47988
2.6		46612	45271	43965	42692	41453	40246	39070	37926	36811	35726
2.7		34670	33642	32641	31667	30720	29798	28901	28028	27179	26354
[illegible]		25551	24771	24012	23274	22557	21860	21182	20524	1988[illegible]	19262
.9		18658	18071	17502	16948	16411	15889	15382	14890	14412	13949
.0		13499	13062	12639	12228	11829	11442	11067	10703	10350	10008
.1	0.0^3	96760	93544	90426	87403	84474	81635	78885	76219	73638	71136
.2		68714	66367	64095	61895	59765	57703	55706	53774	51904	50094
.3		48342	46648	45009	43423	41889	40406	38971	37584	36243	34946
.4		33693	32481	31311	30179	29086	28029	27009	26023	25071	24151
.5		23263	22405	21577	20778	20006	19262	18543	17849	17180	16534
.6		15911	15310	14730	14171	13632	13112	12611	12128	11662	11213
.7		10780	10363	*99611	95740	92010	88417	84957	81624	78414	75324
.8	0.0^4	72348	69483	66726	64072	61517	59059	56694	54418	52228	5012[illegible]
9		48096	46148	44274	42473	40741	39076	37475	35936	34458	33037

Annex 2

Table of Random Numbers

63	47	43	73	86	36	96	47	36	61	46	98	63	71	62	33	26	16	80	45
97	74	24	67	62	43	81	14	57	20	42	53	32	37	32	27	07	36	07	51
[illegible]	76	63	27	66	56	50	26	71	07	31	90	79	78	53	13	55	38	58	59
12	56	85	99	26	96	96	68	27	31	05	03	72	93	15	57	12	10	14	21
55	59	56	35	64	38	54	82	46	22	31	62	43	09	90	06	18	44	32	53
16	22	77	94	39	44	59	43	54	82	17	37	93	23	78	87	35	20	96	43
84	42	17	53	31	57	24	55	96	88	74	04	74	47	67	21	76	33	50	25
63	01	63	78	59	16	95	55	67	19	98	10	50	71	75	12	86	73	58	07
33	21	12	34	29	78	64	56	07	82	52	42	07	14	38	15	53	00	13	42
57	[illegible]	86	32	44	09	44	27	96	34	49	17	46	09	62	90	52	84	77	27
18	18	07	92	46	44	[illegible]	16	58	09	79	83	36	19	62	06	76	50	03	10
[illegible]	62	23	47	75	31	16	07	44	94	83	11	46	37	24	20	14	85	88	45
23	42	40	14	77	82	97	77	77	81	07	45	32	14	08	32	98	94	07	72
62	36	27	19	95	60	39	50	08	20	00	56	76	31	38	80	22	02	53	53
35	87	94	35	12	50	92	26	11	97	42	34	07	96	88	54	41	06	87	93
70	29	17	90	53	13	33	20	38	26	13	89	04	03	77	17	76	37	13	04
56	62	13	37	35	96	85	50	87	75	97	12	25	93	47	70	33	24	03	54
99	19	57	22	77	88	42	95	45	72	16	64	36	16	00	04	43	18	66	79
16	08	15	04	72	33	27	14	34	09	45	59	34	68	49	12	72	07	34	45
[illegible]	[illegible]	73	32	43	50	27	89	87	19	20	15	37	08	49	52	85	66	96	11
18	34	30	12	70	55	74	30	77	40	44	22	78	84	28	04	33	46	09	52
71	57	25	65	75	59	29	97	68	60	71	91	38	67	54	13	58	18	24	76
24	42	37	86	53	48	54	90	65	72	96	57	69	36	10	96	46	92	42	45
00	39	68	29	61	66	37	[illegible]	29	30	77	24	57	03	29	10	45	65	04	26
59	94	93	94	24	68	49	69	10	82	53	75	93	91	30	34	25	20	57	27
15	90	82	66	59	33	62	64	11	12	57	19	00	71	74	60	47	21	29	68
[illegible]	27	74	75	06	06	09	19	74	66	03	94	37	34	02	76	70	90	30	36
35	74	10	16	20	23	32	71	26	30	79	72	43	04	91	16	92	53	56	[illegible]
38	73	16	86	38	41	38	96	01	50	87	75	66	81	41	40	01	74	91	[illegible]
31	94	25	91	47	96	44	33	49	13	34	86	82	53	91	00	52	43	48	[illegible]
66	67	40	67	14	64	05	71	95	86	11	05	65	09	67	75	83	20	33	[illegible]
14	90	54	45	11	75	73	88	05	90	52	27	31	14	36	26	98	12	22	[illegible]
38	65	51	13	60	33	96	02	75	19	07	60	62	93	55	59	33	82	43	[illegible]
29	43	78	73	90	97	51	40	14	02	04	02	33	31	08	39	54	16	49	[illegible]
64	[illegible]	58	[illegible]	[illegible]	15	06	15	93	20	01	90	10	73	06	40	78	78	39	[illegible]
[illegible]	25	97	70	60	22	35	85	15	[illegible]	92	03	51	39	77	59	56	78	08	[illegible]
67	[illegible]	19	88	23	09	98	42	99	04	01	71	62	99	13	63	51	29	16	[illegible]
68	[illegible]	86	85	62	54	87	67	47	51	73	32	08	11	12	44	95	92	73	[illegible]
26	90	01	55	53	58	37	78	80	70	42	10	50	67	42	31	17	33	85	[illegible]
[illegible]	[illegible]	52	[illegible]	73	87	50	36	22	[illegible]	36	78	63	76	35	13	63	76	91	[illegible]